How Strength of Habits Developed as Staff Influences Senior Auditors’ Evaluation of Assumptions Underlying an Estimate

Sarah Bonner
University of Southern California

Kathryn Kadous
Emory University

Tracie Majors
University of Southern California

August 2019

We thank Lori Bhaskar, Scott Emett, Cassandra Estep, Emily Griffith, Bright Hong, Eldar Maksymov, Chris Wolfe, Wendy Wood, and workshop participants at Arizona State University and Texas A&M University for helpful feedback. We thank Erik Fujinami, Bright Hong, JungKoo Kang, Andrew Kim, Sharon Kim, Suteera Pongtepupatham, Taylor Reis, Michael Rezzo, Stacey Ritter, Meredith Schultz, and Fiona Wang for helpful feedback on the instrument. We also thank Nathaniel Young for the photographs used in the instruments, and Bright Hong, Ryan Kalaf, Molly Starobin, Genna Young, and Dana Zadeh for other research assistance. We also thank the four senior auditors with whom we conducted extensive interviews for giving their time. This study was supported by a Center for Audit Quality Research Advisory Board grant. We thank Margot Cella and Lauren Tuite of the CAQ for facilitating access to participants, members of the Research Advisory Board, especially John DeMelis, for helpful input, and participants for giving their time.

The views expressed in this article and its content are those of the authors alone and not those of the Center for Audit Quality.
ABSTRACT

Effective auditing of estimates is critical for audit quality. We posit that audit staff develop and carry over into senior-hood habits to use superficial, piecemeal, and confirmatory cognitive processes, which are effective for staff tasks but ineffective for auditing estimates. Behaviors become habits when they are repeated and concurrently rewarded in a particular context; habits are automatically, unconsciously activated by context cues. In an experiment, we measure audit seniors’ habit strength and manipulate (imagined) context. In a typical audit room context, seniors with stronger habits identify fewer issues with an estimate than those with weaker habits. Seniors with stronger habits perform better in an audit room lacking key context cues and therefore preventing habit activation; seniors with weaker habits do not. This interaction and additional analyses provide strong support for habits as the causal construct. Our findings provide new directions for research on audit quality and implications for audit practice.

JEL codes: G10, M40, M41, M42, D80, D91

Keywords: habits, cognitive processing, accounting estimates, audit quality, impairment, fair value, professional skepticism
I. INTRODUCTION

Accounting estimates, including goodwill and valuation allowances, are based on subjective assumptions made by management and are developed under conditions of measurement uncertainty. These properties make estimates susceptible to misstatement due to management bias, implying that effective audits of estimates are critical for financial reporting quality (e.g., Martin, Rich, and Wilks 2006; Bratten et al. 2013; Griffith, Hammersley, and Kadous 2015; Cannon and Bedard 2017). However, regulators have voiced strong concerns that audits of estimates are ineffective (PCAOB 2012, 2014, 2015, 2017), and inspection reports cite continued deficiencies in auditors’ evaluation of the assumptions underlying estimates (Griffith et al. 2015a; Griffith, Hammersley, Kadous, and Young 2015). Auditors also report difficulty with this task (Griffith et al. 2015a; Glover, Taylor, and Wu 2017; Cannon and Bedard 2017).

In this study, we examine whether audit seniors’ effectiveness at evaluating assumptions underlying an estimate is negatively affected by the strength of cognitive processing habits they develop as staff auditors and carry into senior-hood. A habit is a learned association between a specific behavior and a particular context in which that behavior is performed that develops as the behavior is repeated in the context with concurrent rewards (Wood and Rünger 2016; Mazar and Wood 2018). For example, moviegoers can develop habits to eat popcorn (the behavior) at the theater (the context) if they eat and enjoy popcorn frequently when they see movies in a theater (Neal, Wood, Wu, and Kurlander 2011). Habits vary in strength based on the extent of rewarded repetition in the context. Likewise, we propose that staff auditors can develop habits to engage in specific cognitive processes. These habits will vary in strength based on the extent to which the processes are repeated and rewarded in the typical context in which staff work – an audit room at a client site. Because the processes are not suited for the evaluation of assumptions
underlying an estimate, we predict that seniors who carry into senior-ness stronger “staff habits” will be less effective at evaluating an estimate than will seniors with weaker staff habits.

The specific cognitive processes (labeled “staff processes”) of interest are superficial processing (focusing on obvious, surface features (Craik and Lockhart 1972)), confirmatory processing (searching for and interpreting evidence consistent with expectations or desired outcomes (Kunda 1990; Nickerson 1998)), and piecemeal processing (evaluating cues one or a few at a time, then concluding and moving on (Anderson 1981)). These processes lead to sufficiently effective performance for most tasks performed by staff, such as comparing prices on sales invoices to those on a price list (Westermann, Bedard, and Earley 2015), because these tasks tend to be low in complexity. That is, they require few steps, and each step contains a small set of unambiguous information cues that can be processed individually (Bonner 1994). These tasks also often are performed separately for each item in a sample, and can be viewed as a series of “micro-tasks” for which any necessary integration across items is done later by software or a superior. Moreover, the staff processes are also more efficient than alternative processes, making them particularly attractive when efficiency is emphasized.

Because these processes are both efficient and effective for staff tasks, we propose that some staff will use them and be rewarded for doing so. Because staff generally work in an audit room at a client site, these staff processes can become habitual behaviors (“staff habits”), that is, mentally associated with the context in which they are performed (Wood and Rünger 2016). Like other links in memory, habits (i.e., context-behavior links) vary in strength depending on frequency of co-occurrence of the items (Deese 1960), that is the extent of repetition of the behavior in the context. The stronger the link in memory, the higher the likelihood that the second item (the behavior) is activated when the first item (the context) is experienced (Collins
We hypothesize that, when context cues are present, seniors who have carried over stronger staff habits are more likely to have their habits automatically activated, and thus are more likely to use the staff processes for the task at hand, regardless of task demands, relative to seniors with weaker staff habits. Based on interviews with four seniors, we expect that the “typical” context incorporates both physical (conference room furniture, technology, snacks, drinks, and supplies) and social (evidence of team members in the room) context cues.

The fact that stronger staff habits are automatically activated by context is problematic for auditors who carry them into senior-hood because these seniors likely continue to work in the same context, but now must perform higher complexity tasks that require different cognitive processes. In particular, effective performance of the complex task of evaluating assumptions underlying an estimate requires deep, nonconfirmatory, and integrative processing (e.g., Griffith et al. 2015b). Thus, our first prediction is that, within the typical audit room context, seniors with stronger staff habits will perform worse than seniors with weaker habits, identifying fewer embedded issues with an estimate’s assumptions.

Removing context cues prevents the automatic activation of habitual responses (Wood, Tam, and Witt 2005; Neal et al. 2011). Thus, our second prediction is that seniors with stronger staff habits will identify more issues when they are placed in a context that removes several of the relevant context cues (the “alternative” context). The alternative context should allow these seniors to use processes that are better suited for the estimates task. Finally, because seniors with weaker staff habits are less likely to have the staff processes automatically activated by the typical context, we expect they will benefit less from the alternative context.

We test our predictions in a 2 x 2 between-participants experiment with 128 experienced audit seniors from two large audit firms. Participants evaluate assumptions underlying a fair
value estimate in a goodwill impairment case adapted from Kadous and Zhou (2019). The case contains seven embedded issues reflective of management bias. Each issue requires deep, nonconfirmatory, and integrative processing to identify. Our dependent measure is the number of embedded issues auditors identify.

We measure our first independent variable (staff habit strength) with a reaction time-based measure. The measure captures how fast auditors complete word fragments related to the staff processes after being primed with a photograph of the typical context, relative to how fast they complete non-audit fragments. The baseline allows us to control for reading and typing speed. We use a median split to classify auditors as having stronger or weaker staff habits. Our approach is the most direct method of measuring habits (Rebar, Gardner, Rhodes, and Verplanken 2018) as it captures the strength of the link between the typical context and staff processes in memory (i.e., the definition of habit strength). It also reflects that habitual behavior is largely unconscious, rendering accurate self-reporting difficult (Mazar and Wood 2018).

We manipulate context by randomly assigning auditors to the typical or alternative context. Participants in the typical context view a photograph of a typical audit room (with the cues described earlier) and those in the alternative context view a photograph of the same audit room, but with several physical and social cues removed. All participants then are asked to write a paragraph about how their day would go working in this room, and to imagine themselves working in this room while completing the estimates task.

As predicted, we find that: (1) in the typical context, seniors who have carried over stronger staff habits identify fewer embedded issues than those with weaker habits; (2) seniors with stronger habits identify more issues in the alternative than typical context; and (3) the alternative context results in better performance for seniors with stronger habits but not those
with weaker habits.¹ Path analyses show that, for each hypothesis, the number of identified issues affects seniors’ assessments of reasonableness, which affects whether they take action.

The observed interaction between context and habit strength supports the inference that habit strength drives the observed results. Nonetheless, because the habit construct is new to the accounting literature, we perform additional analyses to strengthen this inference. First, we show that the alternative context does not improve the performance of those seniors with stronger staff habits who likely developed habits in both contexts. Second, consistent with habitual processing operating rapidly (Wood et al. 2014), we show that seniors with stronger habits work faster in the typical than in the alternative context. Third, the alternative context is less helpful to seniors with stronger habits who inhibit the habitual response by exercising self-discipline.

We identify a new factor—*strength of staff cognitive processing habits*—that provides a potential explanation for why deficits in skepticism persist in audits of estimates and other complex accounts despite regulator admonishments and firm tools such as judgment frameworks (PCAOB 2012, 2014, 2015, 2017; Nolder and Kadous 2018). Our insight that ongoing performance deficits can arise from strong habits has important implications for addressing these deficits. Because habits are “hard wired” in memory and activated by context, conventional approaches to improvement, such as training and imposing accountability, are likely ineffective for seniors with strong habits. As science and our practical experiences tell us, it is difficult to overcome strong habits without changing context and/or developing new habits for more appropriate behaviors (Wood 2019). Overcoming habitual after-dinner snacking in the television room, for example, may involve changing the context (e.g., changing rooms, removing the

---

¹ Differences across contexts also could affect auditors’ inferences about other factors (e.g., stress, interruptions). We measure these inferences and conduct tests that rule out these inferences driving our results.
television), and/or repeating a new, rewarded behavior in a particular context in its place (e.g., evening walks with a friend, leading to fresh air and companionship).

Wholesale changes to the typical audit room context are likely infeasible, but small changes that target key context cues may be effective (Wood 2019). For example, seniors could schedule complex tasks when working alone. Firms also could guide seniors to replace old behaviors with new in ways that encourage development of more suitable processing habits. For example, judgment frameworks that encourage the use of complex processes may be more effective if seniors repeatedly use them and are rewarded for proper use. An alternative solution is to prevent strong habits from developing in the first place. In the example above, this might involve frequently restricting evening eating to the kitchen. Audit firms can likewise prevent the development of strong staff habits by assigning staff to assist with “senior tasks” to decrease repetition of the staff processes.

Our findings contribute to the auditing, accounting, and psychology literatures. In studying habits, we add to a growing literature on auditors’ “Type 1” processing (e.g., Wolfe, Christensen, and Vandervelde 2019), and we respond to broader calls for research on the effects of “Type 1” processing on auditor judgments (Griffith, Kadous, and Young 2016). We also extend the literature on auditor skepticism (see Nelson 2009; Nolder and Kadous 2018) by showing that “non-skeptical” processes can become habitual, thus identifying a new antecedent to skepticism. In addition, we develop a measure of auditor staff habit strength that future researchers can use and adapt. We expect that the habits construct and our measurement technique could advance research on the decision making of financial analysts and managers, among others. Finally, we contribute to the habits literature in psychology (Wood 2017), which focuses on physical behaviors; we extend this work to cognitive behaviors.
The rest of the paper is organized as follows. Section 2 provides theory and hypotheses. Sections 3 and 4 describe the design and results of the experiment. Section 5 concludes.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

In this section, we develop theory that superficial, confirmatory, and piecemeal cognitive processes (staff processes) allow for sufficiently effective and particularly efficient completion of tasks that staff typically perform. Drawing from psychology theory on habits, we argue that rewarded repetition of these processes in the typical audit room context can lead staff to develop, and carry over into senior-hood, staff habits of varying strength. For seniors who have carried over stronger staff habits, the staff processes are automatically activated by cues in the typical audit room, regardless of task demands. We then provide evidence that effectively evaluating the assumptions underlying estimates requires cognitive processes that are the opposite of the staff processes. Thus, we predict that seniors with stronger staff habits will be less effective at this task. We further predict that working in an alternative context will improve these auditors’ task effectiveness, but will be less beneficial for seniors with weaker staff habits.

Staff Tasks and Cognitive Processes

Staff auditors typically perform tasks that are low in complexity, such as vouching terms on invoices to supporting documents, confirming cash balances, and inspecting invoices for evidence that credit was authorized (e.g., Abdolmohammadi 1991; Power 2003).² Vouching prices on sales invoices to a price list, for example, is low in complexity because it involves one step and a small number of information cues (Bonner 1994). Further, the cues are unambiguous (i.e., prices are quantified and objective) and can be processed individually (e.g., the price for

² For brevity, we focus on one example of a staff task when explaining why the staff processes allow for effective and efficient performance of staff tasks. However, we expect that the processes are similarly suitable for most staff tasks because these tasks generally have similar characteristics. Seniors in our study reported having spent, on average, 68 percent of their time as staff working on “staff tasks” such as those listed.
one item on an invoice, then for the next), without a need for integration across items, which typically is performed by software or a superior. Based on the processing demands of staff tasks, then, we propose that staff auditors tend to engage in superficial, confirmatory, and piecemeal processing to complete these tasks because they are sufficient for effective task performance and also are particularly efficient, relative to alternative processes that staff may select.

Psychology research defines superficial processing as focusing on obvious, surface features of information cues, and deeper processing as analyzing meaning, inference, and implication of cues (Craik and Lockhart 1972; Craik 2002). The unambiguous cues in staff tasks generally require only superficial processing for effective task performance. For example, the vouching task requires only a quick glance at the price list to see if prices match those on the invoice. Superficial processing also is more efficient than considering evidence in greater depth.

The relatively unambiguous nature of cues in staff tasks and distinct steps for each item being audited also can lead to confirmatory processing, which involves searching for evidence supporting, and/or interpreting information in line with, what one expects or desires to see (Kunda 1990; Nickerson 1998). Again, the nature of the cues requires only confirmatory processing for effective performance. For example, for the vouching task, staff typically see that the price in the price list is correct. Identifying instances in which the price is incorrect does not require nonconfirmatory processing (searching for and accurately evaluating contradictory information) but rather, simply noting the discrepancy. Confirmatory processing also is efficient as it does not require extensive information search or interpretation.

Finally, characteristics of staff tasks also can lead to piecemeal processing, as opposed to integrative processing (Anderson 1981; Garner 1981; Adaval and Wyer 1998). People engaging in piecemeal processing take one or a few cues into working memory, evaluate them, then “close
out” the cues by removing them from working memory. Staff tasks, such as vouching, that consist of a series of “micro-tasks” can be performed effectively with piecemeal processing. That is, since integration is performed later by a competent source (the computer or a senior), effectiveness is not compromised. Piecemeal processing also is efficient because it is quicker than integrative processing, which requires returning to earlier cues for re-processing to consider their implications jointly with those of later cues.

To summarize, the low complexity of staff tasks implies that superficial, confirmatory, and piecemeal processing are sufficiently effective for such tasks and also efficient; consequently, staff likely repeat them. As we describe next, repetition of behavior in a particular context with concurrent rewards leads to habit formation (Mazar and Wood 2018).

**Staff Cognitive Processes as Habits**

A habit is a learned association between a behavioral response and aspects of a particular performance context (Wood and Rünger 2016). The behavior may initially be in response to a goal. For example, a person initially may eat popcorn at a movie theater with the goal of satisfying hunger. However, over time, with rewarded repetition in a particular context, these behaviors can become automatically and unconsciously activated by cues of the context rather than by a person’s situational goal (Wood and Neal 2007). Thus, moviegoers who repeatedly eat and enjoy popcorn can develop an “eat popcorn at the movies” habit that is activated by theater context cues, such as buttery smell, theater seats, and movie previews, rather than by hunger (Wood, Quinn, and Kashy 2002; Wood 2017).

*Habit strength* is the strength of the association in memory between the context and the behavior, with stronger habits resulting from more rewarded repetition of the behavior in the particular context. This phenomenon is best understood by recognizing that habitual behaviors,
even physical ones, are represented in memory as links between the context and behavior (see Wood and Rünger 2016). Links between items in memory, e.g., the concepts “birthday” and “cake,” form as those items are experienced together; strength depends on frequency of co-occurrence (Deese 1960). Thus, in the case of habits, as the frequency with which a person experiences context cues in conjunction with a behavior increases – because he or she repeats the behavior in the context – so does habit strength. When items are linked in memory, the probability that the second item (e.g., the word “cake,” the behavior of eating popcorn) is activated in memory after a person experiences the first item (e.g., the word “birthday,” a movie theater cue) increases with the strength of the link (Collins and Loftus 1975; McNamara 1992). Thus, when context cues are experienced, habit strength determines the probability that habitual behaviors are automatically and unconsciously activated in memory, then enacted irrespective of current goals or task demands. For example, a moviegoer who has a strong habit of eating popcorn at the movies likely will eat popcorn even if he or she recently ate dinner.

While the habits literature has focused on physical behaviors, we believe that the habit construct can be extended to cognitive behaviors where there is a shorter path from the context to the behavior. For physical behaviors such as exercising, context cues activate the cognitive impulse to exercise (Neal, Wood, Labrecque, and Lally 2012), but if a person is busy or tired, this impulse may not lead to the physical act.\(^3\) In contrast, cognitive behaviors can be invoked automatically and effortlessly (Bargh 1994), such that they may be more likely to become habitual. We expect that staff auditors develop habits that vary in strength based on the amount of repetition of and concurrent rewards for staff processes in a particular context. That context is

\(^3\) Consistent with the cognitive element being habitual, Neal et al. (2012) show that people with strong running habits, after being primed with the typical context in which they run, are quicker to identify running-related words than participants with weaker running habits.
the typical audit room at a client site with its attendant physical (conference room and furniture, technology, snacks, drinks, and supplies) and social cues (crowding and clutter that evidence other team members in the room). Repetition will vary based on factors such as frequency of performing staff tasks. Rewards will vary based on factors such as how much satisfaction or praise staff receive for efficient task completion. To summarize, we define staff habit strength as the probability that the typical audit room context automatically activates seniors’ use of the staff processes. We posit that seniors with stronger habits will be less effective at evaluating the assumptions underlying an estimate because it is a complex task that requires different processes.

**Effects of Staff Habit Strength on Seniors’ Evaluation of Assumptions**

Prior research, as well as our task analysis (summarized next), show that evaluating the assumptions underlying an estimate is a complex task that requires deep, nonconfirmatory, and integrative processing. These processes are required because assumptions involve subjective information cues, including predictions; qualitative cues; and cues with implications for other cues, including cues that support one assumption but contradict others (Griffith et al. 2015a, b; Kadous and Zhou 2019; Nolder and Kadous 2018). Interpreting subjective cues requires deep processing; superficial processing could lead to missing evidence indicative of problems. Nonconfirmatory processing is necessary for identifying contradictory evidence and taking it seriously (rather than “explaining it away”). Integrative processing also is necessary for this task; a piecemeal approach could lead to focusing on cues in isolation, missing joint implications.

Research focused on audits of estimates supports this idea. Priming intrinsic motivation (Kadous and Zhou 2019), conveying that an estimate has a high misstatement risk (Griffith 2018), and taking the specialist’s perspective (Joe, Wu, and Zimmerman 2017) lead auditors to better identify an unreasonable estimate by identifying more embedded cues that require deep
and integrative processing. Auditors high in Need for Cognition, which is associated with deep and integrative processing, also are more likely to identify a biased estimate as unreasonable (Griffith et al. 2019). In addition, priming a deliberative mindset (Griffith et al. 2015b), requiring auditors to document contradictory issues (Austin, Hammersley, and Ricci 2019), and asking auditors to consider how management reached their assumptions (Backof, Carpenter, and Thayer 2018) improve consideration of contradictory evidence, leading to lower reasonableness assessments for a biased estimate. Finally, Wolfe et al. (2019) find that asking experienced seniors to think analytically (versus intuitively) when impairment indicator orientation is noted leads them to focus on riskier indicators, and thus assess potential impairment as more likely. Because the staff processes (superficial, confirmatory, and piecemeal) are the opposite of those needed for effective performance of complex tasks, we posit that seniors who carry over stronger staff habits will evaluate the assumptions underlying an estimate less effectively, identifying fewer issues. To reiterate, this will occur because seniors with stronger staff habits have stronger context-response links in memory. These auditors, when exposed to the typical audit room context, will be more likely to have their staff habits automatically activated, and thus they will be more likely to use the staff processes even if the task requires different processes. For these auditors, appropriate processes will be are less accessible, as selecting them requires effortful inhibition of the habitual response. Auditors with weaker habits have weaker links in memory, so staff processes are less likely to be activated in response to context, and, if activated, they can be inhibited less effortfully. For these auditors, alternate (suitable) processes are thus easier to access (Wood, Labrecque, Lin, and Rünger 2014). Stated formally:

**H1:** In the typical audit room context, auditors with stronger staff habits will identify fewer issues in assumptions underlying an estimate than those with weaker staff habits.
Effects of Context on Seniors’ Evaluation of Assumptions

Because staff habits are activated by physical and social cues in the typical context, removing these cues should prevent the automatic activation of the habitual behavior. That is, a change in context can “jar” people out of engaging in habitual behaviors since the original context is not available to activate the response (Wood and Neal 2009). This means that auditors with stronger staff habits, when exposed to an alternative context, should be less prone to use staff processes. Instead, their processing is more likely to be guided by task demands.

Psychology research supports this idea. Wood et al. (2005) show that transfer students maintain their exercise frequency habits only if they continue exercising in the same context in their new school; if they exercise in an alternative context, their exercise frequency changes. Neal et al. (2011) show that people with strong habits to eat popcorn at the movies eat stale popcorn when watching a movie in a theater; when placed in an alternative context (watching a music video in a conference room), they eat popcorn less automatically (when it is fresh but not when it is stale). Using staff cognitive processes to audit an estimate can be likened to eating stale popcorn – it is not what one would do in the absence of habits. We predict that removing key context cues from the typical audit context will “jar” seniors with stronger staff habits out of using the staff processes and, thus, improve their performance. Stated formally:

\[ H2: \] Auditors with stronger staff habits will identify more issues in assumptions underlying an estimate in the alternative audit room context than in the typical audit room context.

Because seniors with weaker staff habits are less likely to have the staff processes activated by cues in the typical context, we expect them to benefit less from being placed in an alternative context. This prediction is consistent with Neal et al. (2011)’s finding that the
behavior of participants with weaker popcorn-eating habits changes less when they are placed in an alternative context. Stated formally:

\[ H3: \] The beneficial effect of the alternative versus typical context will be less pronounced for auditors with weaker staff habits than for those with stronger staff habits.

III. METHOD

Participants

Participants are 128 experienced senior auditors (mean experience of 40.3 months) from two large audit firms.\(^4\) Study administration was facilitated by the Center for Audit Quality, with participants completing the study during firm training sessions.\(^5\) The estimates task requires participants to evaluate management’s assumptions related to a goodwill impairment test. Seniors typically perform this task in practice (Griffith et al. 2015a).

Design

Our study employs a 2 x 2 between-participants design. We measure staff habit strength and use a median split to classify participants as having either stronger or weaker staff habits.\(^6\) We manipulate context by randomly assigning auditors either to the typical audit room or an alternative audit room that eliminates many context cues from the typical audit room.

Staff Habit Strength

Consistent with work in psychology (Wood et al. 2011), we measure staff habit strength using a reaction-time measure of the strength of the link between the typical audit room and staff processes in memory. Recall that habit strength is defined as the strength of the association

---

\(^4\) Institutional review board approval was obtained.

\(^5\) Firm has a significant main effect on the dependent variable, but does not interact with our independent variables. Results for hypothesis tests including firm as a covariate remain significant at the same critical levels as reported.

\(^6\) It is necessary to measure rather than manipulate staff habit strength because the development of habits requires repetition in a particular context over time; consistent with this, the vast majority of studies on habits in psychology measure habit strength (for an exception, see Lin, Wood, and Monterosso 2016).
between context and behavior, so this measure most directly captures habit strength (Rebar et al. 2018). Another advantage of this measure is that it does not rely on self reports—because activation of habitual behavior is largely unconscious, participants likely are unable to accurately report when this occurs (Mazar and Wood 2018).\footnote{People generally are unaware of the context cues that activate habitual behaviors and the fact that such activation occurs; moreover, while people may be able to report on physical behaviors (such as eating popcorn), they generally are less able to provide insight into their cognitive behaviors (Mazar and Wood 2018).}

Specifically, we measure the time it takes to complete \textit{word fragments} for words reflective of the staff processes after auditors are primed with an ecologically valid context (see Wood et al. 2011), here the typical audit room. The strength of the link between the context and behavior in memory affects not only the probability, but also the \textit{speed}, of activation of the behavior when the context is activated. Because we expect that most auditors spend the majority of their time in a context like the typical audit room and develop either weaker or stronger habits in this context, we prime all auditors with this context.\footnote{The median percentage of time our participants report spending in a room like the typical audit room is 70 percent.} \textit{Ceteris paribus}, auditors for whom the staff processes are strongly habitual should more quickly complete the related word fragments. Because reaction times can be affected by both habit strength and idiosyncratic factors (such as typing or reading speed), it is standard practice in psychology to measure habit strength as reaction times to baseline words less reaction times to habit-related words (e.g., Neal et al. 2012).\footnote{For example, in examining strength of running habits, Neal et al. (2012) prime participants with the typical context in which they run and then measure reaction times to “running” and “jogging” relative to reaction times to baseline words unrelated to running, such as “investing” and “thermos.”} Thus, we also measure auditors’ completion times for non-audit control word fragments.

We prime the typical context by having all auditors complete a one-minute “spot the difference” exercise involving two photographs: a typical audit room and the same room with five small differences created using Photoshop. We then collect reaction times by having auditors complete the fragments while the context remains activated (the photograph of the
typical room remains displayed). We instruct auditors to type the entire word represented by each fragment as quickly as possible (see Figure 1, Panel A for instructions). We track completion time in milliseconds. The staff process fragments (e.g., SC_N, GL_NCE), shown in Figure 1, Panel B, jointly reflect superficial, confirmatory, and/or piecemeal processing, and were based on interviews with four seniors. The non-audit control fragments (e.g., B_RK; see Panel B) are roughly matched to the staff process words on number of syllables and length. We measure habit strength as auditors’ average time to complete the staff process fragments minus their average time to complete the control fragments. We observe skewness for some words; thus, we transform reaction time for each fragment by taking the reciprocal (see Whelan 2008).\textsuperscript{10} We classify auditors with faster (slower) completion times relative to the median as having stronger (weaker) staff habits.

Finally, we measure habit strength after manipulating context to prevent hypothesis guessing that might occur were it to be measured at the beginning of the study. We do not expect the earlier manipulation to affect the measurement of habit strength because we take this measure close to the end of the study, after auditors’ memory for context-related behaviors in the estimates task has been cleared by answering a number of post-experimental questions. More important, we (re)prime all auditors with the (same) typical audit room context to measure habit strength. Consistent with this, auditors’ assignment to the typical versus alternative context at the beginning of the study is not associated with their measured habit strength (two-tailed $p = 0.662$).

\textsuperscript{10} Reaction times for the staff process words load on one factor (eigenvalue = 5.01) and have a Cronbach’s alpha of 0.89, supporting the idea that they reliably capture a single construct, which we propose is staff habit strength. Reaction times for the control words load on two factors. Multi-dimensionality is to be expected as the terms are not thematically related.
Audit Room Context

Our second independent variable is audit room context. We manipulate context not to examine its effects *per se* but rather to examine the effects of habit strength when habits are activated (H1) versus not activated (H3), as well as the effects of a context change for auditors with stronger habits (H2). We manipulate context by asking auditors, at the beginning of the study, to imagine they are in one of two audit rooms at a client site as depicted by a photograph – either a typical audit room or an alternative room that removes many of the context cues (shown in Panels A and B of Figure 2). As mentioned earlier, we expect the typical audit room to be the context where much of auditors’ work occurs and, thus, where they are likely to develop staff processing habits, be they stronger or weaker. Consequently, this context is meant to activate the habitual processing during the estimates task, while the alternative context is meant to *not* activate habitual processes. That is, we manipulate context to establish that we are capturing the effects of habit strength rather than some other construct.

The typical room contains key physical and social context cues that auditors encounter when working at client sites. Physical cues include a conference room, its furniture (table and chairs), technology (computer, printer, power cords, computer mouses), snacks and drinks (soda, to-go coffee cups, water bottles, fruit, granola bars, candy), and office supplies (pens, notepads, post-it notes). Social cues provide evidence of other team members being present (crowding and clutter). The alternative context maintains a realistic working environment by including necessary furniture (one chair, the conference table) and technology (one computer, the printer), as well as snacks and drinks and office supplies; however, we omit social cues that evidence the

---

11 This methodology of showing photographs and asking participants to imagine themselves in the context is consistent with psychology research (e.g., Neal et al. 2012; Weinstein, Przybylski, and Ryan 2009).
presence of other team members, such as clutter.\textsuperscript{12} We staged the two audit rooms in a conference room of a participating firm’s offices based on themes gleaned from the interviews with four seniors. A research assistant photographed the rooms.

After viewing the pictures, auditors in the typical (alternative) context condition are given the instructions shown in Panel C (D) of Figure 2. They are asked to “describe what you have imagined in 5-7 sentences,” which is intended to facilitate activation of the context in memory and enable spreading activation to related concepts (the staff processes), as well as to make the auditors feel like they are actually in the room (see MacInnis and Price 1987).

\textbf{Task, Dependent Variables, and Other Measures}

\textit{Estimates Task}

The task requires that auditors evaluate management’s assumptions underlying an estimate of goodwill as part of the client’s step-one analysis of an impairment test and is adapted from Kadous and Zhou (2019).\textsuperscript{13} The task includes background information, the step-one analysis, and evidence related to management assumptions. The client uses a discounted cash flow model to estimate the fair value of the reporting unit, which indicates that it passes the impairment test. The task includes three sections with the key assumptions: five-year projections of revenue, operating expenses, and capital expenditures. We embed seven issues indicative of management bias in these sections. Stronger staff habits (i.e., a higher likelihood that auditors will employ superficial, confirmatory, and/or piecemeal cognitive processing while in a typical audit room context) will reduce the likelihood of identifying each issue because, consistent with

\textsuperscript{12} Our participants found the alternative context to be realistic: there is no difference between the two contexts in auditors’ self-reported extent to which they imagined themselves in the audit room or the extent to which they found imagining themselves in the room natural (smallest two-tailed \(p = 0.478\)).

\textsuperscript{13} Companies are required by accounting standards to perform a step-one analysis when it is more likely than not that a business unit’s book value exceeds its fair value; if the fair value exceeds the book value, the company passes step one of this goodwill impairment test and goodwill is not considered impaired.
the real-world task, identification of each issue requires a combination of deep, nonconfirmatory, and integrative processing. A table with a description of each issue is displayed in the Appendix. After auditors finish reading the case (which is provided to them in hard copy), they are provided a link to the Qualtrics program that captures the dependent variable and other measures.

**Primary Dependent Variable and Other Measures**

Because our focus is on cognitive processing, our primary dependent variable is the number of issues a participant identifies (*Issues Identified*). We ask participants to “list any specific concerns” they have about the estimate in performing their evaluation. An author and a doctoral student, both with auditing experience and blind to experimental condition, independently coded participants’ listings for the presence of each embedded issue and reconciled any differences. Raw agreement was 98% and Cohen’s Kappa is 0.92, which is greater than chance ($p < 0.001$). We use the reconciled coding for our dependent variable.

We also ask auditors to assess the reasonableness of the estimate because auditing standards require this assessment (Griffith et al. 2015a; PCAOB 2019). Focusing the task on this ecologically valid assessment helps to ensure that auditors perform the task as they would on the job (e.g., Kadous and Zhou 2018). We measure auditors’ evaluations of the reasonableness of the fair value (*Reasonableness Assessment*) by asking: “Based on your evaluation, how likely is it that the fair value of Augustin’s reporting unit is reasonable?” with 0 as “Not at all Likely” and 10 as “Extremely Likely.” Because auditors must take action to correct material misstatements (Nelson 2009), we ask auditors to select one of four possible actions they would take (*Action*). Consistent with Griffith et al. (2015b), we classify the first two options as “0” as they convey no current action, and the second two options as “1” as they convey immediate action.14 To

---

14 In the first “no current action” option, the auditor concludes that the fair value is reasonable and marks the task as completed for now; in the second, the auditor delays concluding until the manager is on site. The first “immediate
facilitate additional analyses that validate that we capture habit strength, we measure the extent to which auditors report exerting self-discipline while working on the estimates task, the extent to which they report working in rooms similar to our typical and alternative contexts, and time spent on the estimates task. Finally, we measure potential noise variables such as auditors’ general experience, number of goodwill audits, and comfort auditing goodwill.\footnote{Comfort with goodwill and number of goodwill audits are correlated with Staff Habit Strength (two-tailed $p < 0.05$). However, including these measures as covariates does not change inferences from our tests of hypotheses.}

4. RESULTS

Tests of Hypotheses

Our primary dependent variable is the number of embedded issues related to the assumptions that an auditor identifies; this variable ranges from zero to five of the seven total issues.\footnote{We exclude from our analyses data from one participant who identified six issues. This observation is three standard deviations above the mean and is identified as an outlier. Including this observation does not affect inferences from reported tests.} Descriptive statistics are tabulated in Table 1, Panel A. We test hypotheses using an Analysis of Variance (ANOVA) model with Issues Identified as the dependent variable and independent variables indicating whether the participant’s Staff Habit Strength is stronger or weaker and whether the assigned Context is typical or alternative (see Panel B).\footnote{The data meet the ANOVA assumptions other than that of no outliers. However, because our dependent variable involves count data, we verify that our results hold with Poisson and negative binomial regression models.}

Our first hypothesis predicts that, within the typical context, auditors with stronger staff habits will identify fewer issues than auditors with weaker staff habits. Results of simple effects analyses are displayed in Panel C of Table 1. In support of H1, there is a negative effect of Staff Habit Strength on Issues Identified (one-tailed $p = 0.030$) in the typical context. In this context,

---

action” option involves immediately calling the manager about potential impairment (but not concluding), and the second involves concluding there is a material misstatement and conveying this to the manager and partner.

---

\footnote{Recall that we classify auditors with faster (slower) net reaction times, relative to the median, as having stronger (weaker) staff habits. Results for hypotheses tests are significant at the same critical levels if we (1) eliminate data from the participant with the median net reaction time, (2) reclassify this participant as having weaker instead of stronger habits, and (3) eliminate observations of the five participants with net reaction times closest to the median.}
seniors who have stronger staff habits identify fewer issues related to the estimate than those with weaker staff habits (means = 1.09 vs. 1.77). H2 predicts that auditors with *stronger staff habits* will identify more issues when in the alternative context versus the typical context. In support of H2, among auditors classified as having high *Staff Habit Strength*, there is a positive simple effect of *Context on Issues Identified* (one-tailed $p = 0.021$) (means = 1.83 vs. 1.09); this improvement leads to performance that is indistinguishable from that of auditors with weaker habits in the typical context (untabulated, two-tailed $p = 0.889$).

H3 predicts that the beneficial effect of the alternative context will be smaller for weaker staff habits auditors. As shown in the ANOVA results in Panel B, the interaction between *Staff Habit Strength* and *Context* is significant (two-tailed $p = 0.018$). The pattern of simple effects is consistent with that predicted by H3: the effect of *Context* for auditors with stronger habits is significant, as described above, while the effect of *Context* for auditors with weaker staff habits is not (two-tailed $p = 0.180$). Collectively, these findings support the inferences that our measure captures habit strength and that habitual staff processes are activated in the typical, but not the alternative, context.

---

19 While the effect of *Context* for weaker staff habits auditors is not significant, the effect is directionally negative. We note that psychology studies tend to focus on how context affects people with stronger habits and, therefore, while our prediction of less beneficial effect of context for these auditors is consistent with theory, it is unclear *ex ante* whether this less beneficial effect will be directionally positive or negative. A Johnson-Neymann regression (untabulated, among these auditors) shows a significant inflection point indicating that declines in performance for weaker habits auditors in response to the alternative context occur among auditors who spend more than 68 percent of their time working in the typical context. This finding suggests that a subset of auditors may have developed habits that facilitate deep, nonconfirmatory, and/or integrative processing, i.e., “stronger senior habits,” from working in the typical context that are not activated in the alternative context, thus impeding their performance vis-à-vis the typical context. In other words, these auditors may become “unmoored” when they are placed in a context other than where they have developed habits that are appropriate for the goodwill task (Wood 2019). Consistent with this idea, the most punctual taxpayers have the most difficulty with tax compliance after a disruption in payment context (Dunning, Monestier, Piñeiro, Rosenblatt, and Tuñón 2017).
Later Stage Effects of Habit Strength on Auditors’ Reasonableness Assessments and Action

Auditors may identify issues while auditing estimates, but to correct material misstatements, they must consequently view the estimate as less reasonable and be willing to take action (Nelson 2009). We examine whether issues identified affect these variables using a structural equations model (Byrne 2016), with *Staff Habit Strength* as the independent variable, *Action* as the dependent variable, and *Issues Identified*, then *Reasonableness Assessment* as sequential mediators. We test whether *Context* is a moderator using the approach from Arbuckle (2016) that compares an unconstrained model to a model in which the link between *Staff Habit Strength* and *Issues Identified* is constrained to be equal across *Context* conditions; moderation is evidenced if the unconstrained model shows better fit. The model results are displayed in Figure 3. The unconstrained model fits the data well. The chi-squared test reveals good fit ($\chi^2(6) = 2.69, p = 0.846$), as do other standard measures. The Comparative Fit Index (CFI) of 1.00 is above the threshold of 0.95 (Hu and Bentler 1999), and the Root Mean Square Error of Approximation (RMSEA) of 0.00 is below the 0.05 threshold (MacCallum et al. 1996). As expected, the unconstrained model shows better fit than the constrained model ($\chi^2(1) = 5.73, p = 0.017$).

As shown in Figure 3, Panel A, and consistent with H1, within the typical context, there is a negative effect of *Staff Habit Strength* on *Issues Identified* (one-tailed $p = 0.028$). This relation is not significant in the alternative context (untabulated), consistent with moderation and the results of the test of H3 (two-tailed $p = 0.128$). In both contexts, more *Issues Identified* leads to a lower *Reasonableness Assessment* (largest one-tailed $p = 0.032$), which leads to a higher likelihood of *Action* (both one-tailed $p < 0.001$). To examine H2, we use *Context* as the independent variable, and classify *Staff Habit Strength* as the moderator (see Panel B). As above, the later stage paths are significant (largest one-tailed $p = 0.005$); importantly, *Context* has a
positive effect on *Issues Identified* for auditors with stronger staff habits (one-tailed $p = 0.015$), but not for those with weaker habits (untabulated, two-tailed $p = 0.193$). We conclude that the effects observed in H1-H3 affect auditors’ reasonableness assessments and their actions.

**Further Validation of Habit Strength Measure**

The *Context x Staff Habit Strength* interaction provides strong evidence that our measure of habit strength is capturing that construct. Nevertheless, this measure is new to the accounting literature, so we conduct a number of additional tests motivated by psychology theory to provide further validation.\(^{20}\) In the following subsections, we describe the theoretical prediction related to each factor that we examine and explain how it supports the validity of our inferences about habit strength. We then explain how we measured each factor, and report results of analyses.

**Validation Using Auditors’ Experience in the Alternative Context**

Psychology theory predicts that people develop habits from repeating behaviors in a particular context with concurrent rewards (Wood and Rünger 2016). We predict and find that the alternative context improves the performance of auditors with stronger staff habits; the underlying theory is that the alternative context does not activate the habitual response like the typical context does. However, some auditors with stronger staff habits also may have significant experience working in contexts like the alternative context, and so could have developed staff habits in *that context* as well. For these auditors, we would expect little benefit of the alternative context. On the other hand, auditors with stronger habits who have less experience working in the alternative context should benefit from the alternative context.

---

\(^{20}\) As mentioned earlier, our reaction-time measure is a direct measure of habit strength, and it also has the advantage of being behavioral rather than self-reported. We also attempted to develop questions for participants to self-report their associations in memory between typical audit room context cues and the staff processes. However, these questions did not show evidence of capturing habit strength, likely because people have difficulty self-reporting the strength of cognitive associations.
We measure experience in the alternative context (*Alternative Experience*) for stronger habits auditors based on whether they are below or above the median of 20 percent of time spent in the alternative context, based on self-reported percentage of time spent in this type of audit room. Table 2, Panel A shows descriptive statistics. We examine this prediction using an ANOVA with *Context, Alternative Experience*, and the interaction term, and *Issues Identified* as the dependent variable, for stronger staff habits auditors (see Panel B). The interaction is significant (two-tailed $p = 0.017$); simple effects (Panel C) indicate that *Context* is insignificant for stronger habits auditors who work more in the alternative context (one-tailed $p = 0.373$), but is positive for auditors who work less in this context (one-tailed $p = 0.001$).

**Validation Using Auditors’ Speed of Reviewing the Goodwill Case**

Psychology theory predicts that, if a behavior is strongly habitual, it occurs more rapidly than if that it is not habitual (Wood et al. 2014). Thus, if stronger habits auditors are working habitually in the typical context, we expect they will work at a quicker pace relative to when they are working in the alternative context (and not using habitual processes). We measure working speed using minutes spent reviewing the case materials, as calculated by participants’ recorded start and end times (*Time*). Descriptive statistics are reported in Table 3, Panel A. We examine this prediction using a t-test across *Context* conditions, with *Time* as the dependent variable. Consistent with the prediction, the significant difference in *Time* across *Context* (see Panel B) reveals that they work more quickly in the typical context (one-tailed $p = 0.027$).

---

21 While 20 percent of their time may seem small, the amount of time auditors spend in the typical context doing staff tasks is more than sufficient to develop habits of any strength (Wood 2019). For example, a very conservative estimate using a five-day workweek, two weeks vacation time, two years as a staff, and one staff task per day, provides 100 opportunities for repetition of the staff processes in that context.

22 It is possible that features of the alternative context (e.g., lack of colleagues) could lead auditors to work more slowly in this context. However, weaker staff habits auditors show no significant difference in time across *Context* conditions (untabulated, two-tailed $p = 0.266$), which helps to rule out such an explanation.
Validation Using Auditors’ Self-Discipline Exerted While Working on the Estimates Task

Psychology studies show that strongly habitual behaviors can be avoided by individuals exerting self-discipline to effortfully inhibit the habitual response, or because an alternative context prevents their activation (Neal et al. 2013). Thus, if habit strength is driving stronger staff habits auditors’ lower performance in the typical context (and prevention of activation is driving their improvement in the alternative context), we expect two results related to self-discipline for these auditors. First, within the typical context, performance should improve as they exert more self-discipline (if some attempt to inhibit their habits). Second, the alternative context should have less benefit for stronger habits auditors who exert high self-discipline.

We capture self-discipline using auditors’ ratings of agreement on a 7-point scale that they exerted self-discipline while working on the estimates task. Results of a regression model among stronger staff habits auditors (displayed in Panel A of Table 4) with Context, Self-Discipline, and the interaction term, and Issues Identified as the dependent variable, reveal a significant Context x Self-Discipline interaction (two-tailed $p = 0.045$). The positive effect of Self-Discipline within the typical context (see Panel B) supports that auditors with stronger staff habits who exert more self-discipline perform better. Additional estimates from the model also support that the nature of the interaction is that, as auditors with stronger staff habits exert more self-discipline, they experience less benefit from the alternative context. The alternative context has a significant positive effect for those exerting low levels of self-discipline (one-tailed $p = 0.003$), but no effect for those exerting higher levels of self-discipline (one-tailed $p = 0.497$).23

---

23 One could argue that this pattern of results reflects differences across the typical and alternative contexts, including more distractions in the typical context. Exerting self-discipline could benefit auditors more in the typical context, where there are more distractions to overcome. However, weaker staff habits auditors (untabulated) show no effect of Self-Discipline in the typical context (one-tailed $p = 0.357$) and there is no Context x Self-Discipline interaction (two-tailed $p = 0.791$) for these auditors, which helps to rule out such an explanation.
To summarize, these analyses provide strong support that our causal construct is habit strength. First, the alternative context is less beneficial for stronger habits auditors who have more experience in that context. Because these auditors likely also developed these strong habits in the alternative context, this result supports habits being at play. Second, the observation that auditors with stronger staff habits work more quickly in the typical context supports habitual processing in that context (and non-habitual processing in the alternative context). Third, that stronger habits auditors perform better by either exerting self-discipline in the typical context or, if they exert little self-discipline, through the alternative context’s inhibiting the habitual response, provides further support.

Robustness Checks for Differences Across the Typical and Alternative Contexts

The alternative context removes cues from the typical context to inhibit activation of habitual responses, but these differences also could affect auditors’ inferences, and thus, their cognitive processing. For example, if auditors infer they are in busy season (not in busy season) in the typical (alternative) context, they may be more (less) prone to engage in superficial processing. These differences would create a main effect of context rather than the predicted and observed interaction with habit strength, but we nevertheless examine this issue. Table 5, Panel A reports auditors’ inferences; Chi-square analyses (see Panel B) indicate that three inferences differ across contexts. Auditors are more likely to infer in the typical context that they would be stressed, interrupted, and in busy season. As only five auditors mention busy season, we

---

24 Finally, we attempted to provide additional validation for our measure of habit strength by examining whether a measure of repetition of staff processes (when the seniors were staff) was predictive of habit strength. This measure did not have satisfactory reliability, and we do not report these analyses.

25 We examine this issue by having two research assistants code auditors’ written paragraphs for inferences that they would (1) be reviewing staff work, (2) feel stressed, (3) experience interruptions from team members, (4) be in busy season, and (5) be contacted by the client. The assistants coded whether the auditor mentioned the inference, mentioned the opposite (e.g., that they would not be reviewing work), or did not mention the inference. Cohen’s Kappa is 0.88, which is significantly greater than chance (p < 0.001). Coding differences were reconciled.
confirm that our tests of hypotheses replicate when excluding these five participants. Then, we examine the effects of the stressed and interruption inferences in three ways. First, we include them as covariates in the ANOVAs used to test hypotheses; they are never significant predictors of Issues Identified.\textsuperscript{26} Second, in lieu of using Context as an independent variable in our analyses, we use the inferences. The inferences are not significant when used as a proxy for the effects of context. Third, for these inferences to explain our interaction results, stronger habits auditors would need to show a greater reduction in the inferences across the typical and alternative contexts than weaker habits auditors. ANOVAs using Staff Habit Strength and Context and their interaction, with the inferences as dependent variables, show only a main effect of Context.

V. DISCUSSION AND CONCLUSIONS

Audits of estimates are critical for financial reporting quality, but regulators have voiced concerns that audits of estimates are ineffective, particularly citing continued deficiencies in auditors’ evaluations of the assumptions underlying estimates (PCAOB 2012, 2014, 2015, 2017; Griffith et al. 2015a). We find evidence that audit seniors’ effectiveness at evaluating assumptions underlying an estimate is negatively affected by the strength of cognitive processing habits they develop as staff auditors and carry into senior-hood. These staff habits, while effective for the low complexity tasks that staff perform, are ineffective for more complex tasks, including evaluating assumptions. Using a reaction-time measure of staff habit strength and manipulating imagined context, we show that, in the typical audit room context, seniors with stronger staff habits identify fewer embedded issues with assumptions than do seniors with weaker staff habits. Stronger staff habits auditors placed in an alternative context that removes

\textsuperscript{26} We run these analyses several ways. First, we code auditors who mention the opposite as 0, do not mention as 1, and mention as 2. Second, we create a dichotomous measure of auditors who mention vs. do not (i.e., not mention or mention opposite). Third, we repeat the second analysis excluding auditors not mentioning the inference.
several context cues, thus preventing the automatic activation of habitual processes, identify more embedded issues than they do in the typical context. Identifying more issues leads stronger staff habits auditors to assess the estimate as less reasonable and be more likely to take action.

As predicted, we also observe an interaction between context and habit strength, indicating that auditors with weaker habits benefit less from the alternative context. The interaction provides strong support that habit strength is the causal construct. However, we conduct additional analyses to further strengthen our inferences. First, we find that seniors who likely developed strong habits in both our typical and alternative contexts do not perform better in the alternative context. Second, stronger habits seniors work more quickly in the typical context than the alternative context, which is consistent with habitual processing operating more rapidly than non-habitual processing (Wood et al. 2014). Third, seniors with stronger habits who are already inhibiting the habitual response by exerting self-discipline experience less benefit from working in the alternative context.

In light of regulators’ concerns that professional skepticism is wanting in audits of estimates, our findings suggest that seniors who carry over stronger staff habits may have difficulty exercising skepticism because they continue to work in a context that automatically activates superficial, piecemeal, and/or confirmatory processing, all of which are antithetical to skeptical processing (Griffith et al. 2015b; Nolder and Kadous 2018; PCAOB 2017, 35). Because habits are links in auditors’ memory between the context and these cognitive processes, interventions to mitigate the negative effects of habit strength must either break this cognitive link or prevent it from becoming strong in the first place (Verplanken and Wood 2006, 92). Traditional approaches such as regulator admonishments or training are likely ineffective because they do not address the link. For example, the use of more neutral, less confirmatory
words in the PCAOB’s proposed standard on auditing estimates (PCAOB 2017) may be useful in helping weaker habits auditors reframe their task, but may go unnoticed by a senior with stronger habits who is working in the typical audit room context, as the new language does not disrupt the context-behavior link. In addition, while a senior with stronger habits may comply with the PCAOB’s directive to find more contradictory evidence (PCAOB 2017), she might “explain away” this evidence since people with strong habits search for ways to continue their habitual behaviors (Verplanken and Wood 2006). Identifying effective interventions is also important because we expect staff habits to impede seniors’ performance of other tasks that require complex processing, such as revenue recognition (PCAOB 2017).

We propose that audit firms may be able to improve seniors’ performance of complex tasks by directly targeting the context-behavior link. While a wholesale change in context is likely infeasible for audit firms, individual seniors could purposefully allocate time for complex tasks when key context cues such as team members are absent. Such situational self-control strategies (Duckworth, Gendler, and Gross 2016) can be effective, over time, in breaking habits (Wood 2019). Firms could also explore guiding seniors to develop new habits to engage in the more sophisticated processing necessary to perform complex tasks; for example, tools such as judgment frameworks meant to encourage this processing could lead to longer-term behavioral change via new habit formation if used properly in a consistent context and accompanied by rewards. Finally, firms could attempt to prevent the development of strong habits among staff auditors by taking measures that decrease repetition of the staff processes, for example, by assigning both staff tasks and assistance with senior tasks. Alternatively, firms could encourage staff to engage in deep, nonconfirmatory, and integrative processing when doing staff tasks, for
example, through the use of visualization that could induce integrative processing by making salient a broader view of test results.

Our study contributes to the auditing, accounting, and psychology literatures. We demonstrate how habits, a form of Type 1 processing, affect auditor judgments, and thus we contribute to a growing literature in this area (see, e.g., Wolfe et al. (2019) on intuitive processing). In so doing, we respond to broader calls for more research on how “Type 1” processing affects auditor judgments (Griffith et al. 2016). Our work also contributes to the literature on auditor skepticism (see Nelson 2009; Nolder and Kadous 2018) by providing an alternative view of skepticism failures as arising from habits. We provide a methodology for measuring these cognitive processing habits that future auditing studies can incorporate. We expect that the habits construct applies to a wide range of decision making in accounting. For example, like audit staff, junior analysts repeatedly perform low complexity tasks, and then perform more complex tasks as senior analysts. Habits may affect which analysts “survive,” as well as forecast accuracy among those who do survive (e.g., Clement, Koonce, and Lopez 2007; Bradley, Gokkaya, and Liu 2017). Finally, we contribute to the literature in psychology on habits (see Wood 2017) by showing that habits can be cognitive; extant research has focused on physical behaviors.

Our study is subject to limitations that offer opportunities for future research. Consistent with psychology theory on habits, we find that auditors with weaker staff habits benefit less from the alternative context. However, we do not examine whether weaker staff habits auditors are also those with stronger senior habits (i.e., habits to engage in deep, nonconfirmatory, and integrative processing). Future research could adapt our methodology for measuring habits to examine the effects of the strength of senior habits.
References


Panel A:

**Instructions:**
On the following screens, you will see a series of word fragments. The number of blanks indicates the number of missing letters.

For example, if there is one blank visible, this means there is one letter missing. As another example, if there are three blanks visible, this means there are three letters missing.

Please type the word *(the entire word, not just the missing letters)* as quickly as possible and then hit the arrow to continue to the next word fragment. As an example to get you started, view the two word fragments below:

**WO __ D**

**__ __ AGMEN __**

As soon as you figure out the word, you would type the word in the textbox below the word fragment. For example, once you realized the first word was “WORD,” you would type “WORD” in the textbox. As another example, once you realized the second word was "FRAGMENT," you would type "FRAGMENT" in the textbox.

Please remember to type in the word and hit the arrow button as quickly as possible. Please click the arrow below to begin.

Panel B:

<table>
<thead>
<tr>
<th>CONTROL WORD FRAGMENTS</th>
<th>STAFF PROCESS WORD FRAGMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• B ___ RK (BARK)</td>
<td>• CHE ___ LIST (CHECKLIST)</td>
</tr>
<tr>
<td>• ___ HOC ___ ATE (CHOCOLATE)</td>
<td>• GLAN ___ E (GLANCE)</td>
</tr>
<tr>
<td>• FANC ___ (FANCY )</td>
<td>• INSPE ___ (INSPECT)</td>
</tr>
<tr>
<td>• MAGAZ ___ ES (MAGAZINES)</td>
<td>• PROC ___ D (PROCEED)</td>
</tr>
<tr>
<td>• MA ___ SHMA ___ OW (MARSHMALLOW)</td>
<td>• ___ ASONABLE (REASONABLE)</td>
</tr>
<tr>
<td>• PAINT ___ USH (PAINTBRUSH)</td>
<td>• SC ___ N (SCAN)</td>
</tr>
<tr>
<td>• ___ TERMELON (WATERMELON)</td>
<td>• ___ KIM (SKIM)</td>
</tr>
<tr>
<td></td>
<td>• SPA___ ___ EET (SPREADSHEET)</td>
</tr>
<tr>
<td></td>
<td>• VE ___ FY (VERIFY)</td>
</tr>
</tbody>
</table>

Panel A displays the instructions that participants read prior to completing the word fragments. Panel B displays the word fragments that we use to measure *Staff Habit Strength*. We calculate the average of each auditor’s completion times for the staff process fragments minus the average of each auditor’s completion times for the control (non-audit related) word fragments. We transform each reaction time by taking the reciprocal, which is a transformation commonly used in psychology studies using reaction-time measures to adjust for skewness (see Whelan 2008). We then classify auditors with completion times faster (slower) than the median as having stronger (weaker) staff habits.
FIGURE 2
Audit Room Context Manipulation

Panel A: Typical Audit Room Context

Panel B: Alternative Audit Room Context
Panel C: Instructions Read by Participants in the Typical Audit Room Context:

- Please imagine that you are working in this audit room today.
- You are working in here all day, and your intern and three staff are also all here today.
- Look around and take in the room. Imagine how the day would progress as you are working in this room.

For example:
  - Imagine yourself sitting in the chair (yours is the gray one to the far right) and your staff and intern sitting in the other chairs.
  - Imagine yourself using your laptop and other tools/supplies as you are doing your audit work.
  - Imagine what might be happening throughout the day as you do your work in this room.

Please describe what you have imagined in 5-7 sentences in the box below.

Panel D: Instructions Read by Participants in the Alternative Audit Room Context:

- Please imagine that you are working in this audit room today.
- You are working by yourself in here all day, as your intern and three staff are all at a full day training in the local office today.
- Look around and take in the room. Imagine how the day would progress as you are working in this room.

For example:
  - Imagine yourself sitting in the chair.
  - Imagine yourself using your laptop and other tools/supplies as you are doing your audit work.
  - Imagine what might be happening throughout the day as you do your work in this room.

Please describe what you have imagined in 5-7 sentences in the box below.

Figure 2 provides details of the manipulation of (imagined) context. Auditors assigned to the typical (alternative) audit room context view the picture of the audit room in Panel A (Panel B), and read the instructions displayed in Panel C (and Panel D). After reading through the instructions, auditors then write a paragraph imagining themselves working in the context in the picture.
Results of Process Model Examining Effects of Issues Identified on Reasonableness Assessments and Action

**Panel A:**

- **H1 (in the typical context):**
  - **Staff Habit Strength**
  - $\beta = -0.686$
  - $p = 0.028$

- **Issues Identified**
  - $\beta = -0.577$
  - $p < 0.001$

- **Reasonableness Assessment**
  - $\beta = -0.123$
  - $p < 0.001$

- **Action**

**H3 (moderating effect):**

- **Context**

**Panel B:**

- **Issues Identified**
  - $\beta = +0.745$
  - $p = 0.015$

- **Reasonableness Assessment**
  - $\beta = -0.423$
  - $p = 0.005$

- **Action**
  - $\beta = -0.100$
  - $p < 0.001$

---

*Staff Habit Strength* is our measured independent variable, and is defined in the notes to Figure 1. *Context* is our manipulated independent variable, and is defined in the notes to Figure 2. *Issues Identified* is the total number of issues, out of seven embedded issues, that the auditor identifies in the goodwill impairment case. We measure *Reasonableness Assessment* on a 10-point scale, with the following question: “Based on your evaluation, how likely is it that the fair value of Augustin’s reporting unit is reasonable?” with 0 as “Not at all Likely” and 10 as “Extremely Likely.” We measure *Action* based on whether auditors choose, from four possible actions, one of the two that conveys taking immediate action, or one of the two that conveys not taking immediate action. We test the model using a structural equations modeling approach (Byrne 2016). Panel A displays the model within the typical context. The figure shows how *Staff Habit Strength* leads to fewer *Issues Identified* within the typical context (i.e., H1). The chi-squared test (for the overall model, which is across the typical versus alternative contexts) reveals good fit ($\chi^2 \beta = 2.69$, $p = 0.846$), as does the RMSEA of 0.00 and CFI of 1.00. As evidence that *Staff Habit Strength* moderates the effect of *Context* (i.e., H3), the unconstrained model shows better fit than the model constraining the initial link to be equal across *Contexts* ($\chi^2 \alpha = 5.73$, $p = 0.017$). Panel B shows the positive effect of *Context* on *Issues Identified* for auditors with stronger staff habits (i.e., H2). As shown above, in both instances, greater *Issues Identified* leads to lower *Reasonableness Assessment*; lower *Reasonableness Assessments* then lead to greater *Action*. All p-values are one-tailed for directional predictions.
TABLE 1: Auditor Performance (Issues Identified) by Staff Habit Strength and Context

Panel A: Descriptive Statistics

<table>
<thead>
<tr>
<th>Context</th>
<th>Stronger Staff Habit Auditors</th>
<th>Weaker Staff Habit Auditors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Audit Room</td>
<td>1.09 (1.19) n=34</td>
<td>1.77 (1.71) n=31</td>
</tr>
<tr>
<td>Alternative Audit Room</td>
<td>1.83 (1.58) n=30</td>
<td>1.28 (1.30) n=32</td>
</tr>
</tbody>
</table>

Panel B: Analysis of Variance for Issues Identified

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staff Habit Strength</td>
<td>1</td>
<td>0.14</td>
<td>0.14</td>
<td>0.07</td>
<td>0.795</td>
</tr>
<tr>
<td>Context</td>
<td>1</td>
<td>0.50</td>
<td>0.50</td>
<td>0.24</td>
<td>0.626</td>
</tr>
<tr>
<td>Staff Habit Strength x Context (H3)</td>
<td>1</td>
<td>12.14</td>
<td>12.14</td>
<td>5.77</td>
<td>0.018</td>
</tr>
<tr>
<td>Error</td>
<td>123</td>
<td>258.79</td>
<td>2.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C: Simple Effects Comparisons

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context for Stronger Staff Habit Auditor (H2)</td>
<td>123</td>
<td>2.05</td>
<td>0.021*</td>
</tr>
<tr>
<td>Context for Weaker Staff Habit Auditors</td>
<td>123</td>
<td>-1.35</td>
<td>0.180</td>
</tr>
<tr>
<td>Staff Habit Strength in the Typical Context (H1)</td>
<td>123</td>
<td>-1.90</td>
<td>0.030*</td>
</tr>
<tr>
<td>Staff Habit Strength in the Alternative Context</td>
<td>123</td>
<td>1.50</td>
<td>0.137</td>
</tr>
</tbody>
</table>

We conduct an ANOVA to test our hypotheses. Independent variables are defined in the notes to Figures 1 and 2. The dependent variable is Issues Identified, which is the total number of issues, out of seven embedded issues, that the auditor identifies in the goodwill impairment case. Descriptive statistics are reported in Panel A. Panel C reports our test of H1, that is, the simple effect of Staff Habit Strength on Issues Identified, considering the typical audit room context. Panel C also reports our test of H2, which is the simple effect of Context on Issues Identified, considering stronger staff habits auditors. Panel B reports our test of H3, that is, the interactive effect of our independent variables on Issues Identified. P-values with * are one-tailed, and all other p-values are two-tailed.
TABLE 2: Performance (Issues Identified) for Stronger Staff Habits Auditors by Context and Their Experience with the Alternative Context

Panel A: Descriptive Statistics

<table>
<thead>
<tr>
<th>Context</th>
<th>Less Experience with Alternative Context</th>
<th>More Experience with Alternative Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Audit Room</td>
<td>0.94 (1.25) n=17</td>
<td>1.24 (1.15) n=17</td>
</tr>
<tr>
<td>Alternative Audit Room</td>
<td>2.41 (1.70) n=17</td>
<td>1.08 (1.04) n=13</td>
</tr>
</tbody>
</table>

Panel B: Analysis of Variance for Issues Identified

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Experience</td>
<td>1</td>
<td>4.27</td>
<td>4.27</td>
<td>2.44</td>
<td>0.123</td>
</tr>
<tr>
<td>Context</td>
<td>1</td>
<td>6.80</td>
<td>6.80</td>
<td>3.88</td>
<td>0.053</td>
</tr>
<tr>
<td>Alternative Experience x Context</td>
<td>1</td>
<td>10.47</td>
<td>10.47</td>
<td>5.98</td>
<td>0.017</td>
</tr>
</tbody>
</table>

Panel C: Simple Effects Comparisons

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context in less Alternative Experience</td>
<td>60</td>
<td>3.24</td>
<td>0.001*</td>
</tr>
<tr>
<td>Context in more Alternative Experience</td>
<td>60</td>
<td>-0.32</td>
<td>0.373*</td>
</tr>
<tr>
<td>Alternative Experience in the Typical Context</td>
<td>60</td>
<td>0.65</td>
<td>0.519</td>
</tr>
<tr>
<td>Alternative Experience in the Alternative Context</td>
<td>60</td>
<td>-2.74</td>
<td>0.008</td>
</tr>
</tbody>
</table>

We conduct an ANOVA for this additional construct validation analysis among stronger staff habits auditors (*Staff Habit Strength* is defined in the notes to Figure 1). The independent variable of *Context* is defined in the notes to Figure 2. The independent variable of *Alternative Experience* is based on auditors’ self-reported percentage of time spent on their engagements in an audit room like our alternative context. We classify auditors as having less versus more *Alternative Experience* using a median split (at 20 percent of their time spent in rooms like the alternative context). The dependent variable is *Issues Identified*, which is defined in the notes to Table 1. Descriptive statistics are reported in Panel A. The results of the ANOVA are reported in Panel B, and simple effects are reported in Panel C. *P*-values with * are one-tailed, and all other *p*-values are two-tailed.
TABLE 3: Time Spent for Stronger Staff Habits Auditors Across Contexts

Panel A: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Context</th>
<th>Typical Audit Room</th>
<th>Alternative Audit Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stronger Staff Habits Auditors</td>
<td></td>
<td>10.84 (3.78)</td>
<td>13.38 (5.67)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=28</td>
<td>n=28</td>
</tr>
</tbody>
</table>

Panel B: T-test for *Time Spent Reviewing Case*

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>54</td>
<td>1.97</td>
<td>0.027</td>
</tr>
</tbody>
</table>

We conduct a t-test across *Context* conditions for this additional construct validation analysis among strong staff habits auditors (*Staff Habit Strength* is defined in the notes to Figure 1). *Context* is defined in the notes to Figure 2. The dependent variable is *Time Spent*, which is auditors’ self-reported time spent reviewing the goodwill case. As eight stronger staff habits auditors did not report their start time, end time, or both, n is 56. Descriptive statistics are reported in Panel A. The result of the t-test is reported in Panel B. The *p*-value is one-tailed.
**TABLE 4: Performance (Issues Identified) for Stronger Staff Habits Auditors by Context and Self-Discipline While Working on the Case**

**Panel A: Regression Model**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Coeff.</th>
<th>Std. Error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>60</td>
<td>-0.47</td>
<td>0.85</td>
<td>-0.55</td>
<td>0.583</td>
</tr>
<tr>
<td>Context</td>
<td>60</td>
<td>3.61</td>
<td>1.46</td>
<td>2.47</td>
<td>0.016</td>
</tr>
<tr>
<td>Self-Discipline</td>
<td>60</td>
<td>0.32</td>
<td>0.17</td>
<td>1.91</td>
<td>0.061</td>
</tr>
<tr>
<td>Context x Self-Discipline</td>
<td>60</td>
<td>-0.58</td>
<td>0.28</td>
<td>-2.05</td>
<td>0.045</td>
</tr>
</tbody>
</table>

**Panel B: Simple Effects Estimates**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Coeff.</th>
<th>Std. Error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context for Low Self-Discipline Auditors</td>
<td>60</td>
<td>1.47</td>
<td>0.51</td>
<td>2.90</td>
<td>0.003*</td>
</tr>
<tr>
<td>Context for Medium Self-Discipline Auditors</td>
<td>60</td>
<td>0.74</td>
<td>0.34</td>
<td>2.15</td>
<td>0.018*</td>
</tr>
<tr>
<td>Context for High Self-Discipline Auditors</td>
<td>60</td>
<td>0.00</td>
<td>0.48</td>
<td>0.01</td>
<td>0.497*</td>
</tr>
<tr>
<td><strong>Self-Discipline</strong> in Typical Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>0.32</td>
<td>0.17</td>
<td>1.91</td>
<td>0.030*</td>
</tr>
<tr>
<td><strong>Self-Discipline</strong> in Alternative Context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>-0.25</td>
<td>0.23</td>
<td>-1.13</td>
<td>0.265</td>
</tr>
</tbody>
</table>

We conduct a regression analysis for this additional construct validation analysis among auditors with stronger staff habits (*Staff Habit Strength* is defined in the notes to Figure 1). The independent variable of *Context* is defined in the notes to Figure 2. The independent variable of *Self-Discipline* is measured as auditors’ agreement on a 7 point scale that they exerted self-discipline while working on the goodwill task. The dependent variable is *Issues Identified*, which is defined in the notes to Table 1. The results of the regression model are reported in Panel A. Estimates from the model for the effect of *Context* at low (one standard deviation below the mean), medium (the mean), and high (one standard deviation above the mean) levels of self-discipline are reported in Panel B, as well as estimates of the effect of *Self-Discipline* within the Typical and Alternative *Context* conditions. *P*-values with * are one-tailed, and all other *p*-values are two-tailed.
TABLE 5: Inferences Made by Auditors in Coding of Paragraphs (by Context)

Panel A: Descriptive Statistics – Frequencies (Percentages)

<table>
<thead>
<tr>
<th>Inference</th>
<th>Typical Context</th>
<th>Alternative Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewed Staff’s Work</td>
<td>4 (6.2%)</td>
<td>3 (4.8%)</td>
</tr>
<tr>
<td>Stress (i.e., due to clutter, being cramped)</td>
<td>34 (52.3%)</td>
<td>4 (6.5%)</td>
</tr>
<tr>
<td>Interruptions</td>
<td>32 (49.2%)</td>
<td>2 (3.2%)</td>
</tr>
<tr>
<td>Busy Season</td>
<td>5 (7.7%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Contact from the Client</td>
<td>16 (24.6%)</td>
<td>21 (33.9%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inference</th>
<th>Typical Context</th>
<th>Alternative Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Mention of the Inference</td>
<td>61 (93.8%)</td>
<td>58 (93.6%)</td>
</tr>
</tbody>
</table>

Panel B: Chi-Square Test for Differences Across Context Conditions

<table>
<thead>
<tr>
<th>Inference</th>
<th>Chi-Square Test Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewed Staff’s Work</td>
<td>$\chi^2(2) = 1.15, p = 0.563$</td>
</tr>
<tr>
<td>Stress (i.e., due to clutter, being cramped)</td>
<td>$\chi^2(2) = 34.60, p &lt; 0.001$</td>
</tr>
<tr>
<td>Interruptions</td>
<td>$\chi^2(2) = 47.08, p &lt; 0.001$</td>
</tr>
<tr>
<td>Busy Season</td>
<td>$\chi^2(1) = 4.97, p = 0.026$</td>
</tr>
<tr>
<td>Contact from the Client</td>
<td>$\chi^2(1) = 1.32, p = 0.251$</td>
</tr>
</tbody>
</table>

The above table displays our analysis of inferences auditors make by whether they are assigned to the typical or alternative Context condition. Panel A displays frequencies and percentages (coded from auditors’ written paragraphs during the manipulation of context, i.e., in which they imagined themselves working in the particular context). Panel B displays results of a chi-squared test for differences in frequencies across Context conditions.
### APPENDIX – EMBEDDED ISSUES IN GOODWILL IMPAIRMENT CASE

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Description of the issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue projections</td>
<td>The company consistently over-projected growth in the past, which casts doubt on the accuracy of the current projections</td>
</tr>
<tr>
<td>Revenue projections</td>
<td>The projected revenue growth of Product C is not guaranteed due to uncertainties (a new competing product, delays in production, and/or reliance on synergies with existing product lines)</td>
</tr>
<tr>
<td>Revenue projections</td>
<td>There is an outlier in the benchmarking analysis for the projected revenue growth rate, so while the client’s rate is below the peer average, it would not be if this outlier were excluded</td>
</tr>
<tr>
<td>Revenue projections</td>
<td>Projected revenue growth is inconsistent with the overall market/economy/industry outlook</td>
</tr>
<tr>
<td>Operating expense projections</td>
<td>The company plans to increase sales staff by 10 percent in the next three years, resulting in a significant increase in employment expense. This is not factored into the client’s operating expense assumption</td>
</tr>
<tr>
<td>Capital expenditures projections</td>
<td>The company is building a new $14 million office building, which is not included in the capital expenditures forecast.</td>
</tr>
<tr>
<td>Capital expenditures projections</td>
<td>The company’s forecasted capital expenditures exhibit slower growth than industry analysts’ projections.</td>
</tr>
</tbody>
</table>